



## Characterization of CA-MRSA TCH1516 exposed to nafcillin in bacteriological and physiological media

**Poudel, Saugat; Tsunemoto, Hannah; Meehan, Michael; Szubin, Richard; Olson, Connor A.; Lamsa, Anne; Seif, Yara; Dillon, Nicholas; Vrbanc, Alison; Sugie, Joseph**

*Total number of authors:*  
18

*Published in:*  
Scientific Data

*Link to article, DOI:*  
[10.1038/s41597-019-0051-4](https://doi.org/10.1038/s41597-019-0051-4)

*Publication date:*  
2019

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

### *Citation (APA):*

Poudel, S., Tsunemoto, H., Meehan, M., Szubin, R., Olson, C. A., Lamsa, A., Seif, Y., Dillon, N., Vrbanc, A., Sugie, J., Dahesh, S., Monk, J. M., Dorrestein, P. C., Pogliano, J., Knight, R., Nizet, V., Palsson, B. O., & Feist, A. M. (2019). Characterization of CA-MRSA TCH1516 exposed to nafcillin in bacteriological and physiological media. *Scientific Data*, 6(1), [43]. <https://doi.org/10.1038/s41597-019-0051-4>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

OPEN

DATA DESCRIPTOR

# Characterization of CA-MRSA TCH1516 exposed to nafcillin in bacteriological and physiological media

Saugat Poudel<sup>1</sup>, Hannah Tsunemoto<sup>2</sup>, Michael Meehan<sup>3,4</sup>, Richard Szubin<sup>1</sup>, Connor A. Olson<sup>1</sup>, Anne Lamsa<sup>2</sup>, Yara Seif<sup>1</sup>, Nicholas Dillon<sup>5</sup>, Alison Vrbanc<sup>5</sup>, Joseph Sugie<sup>2</sup>, Samira Dahesh<sup>5</sup>, Jonathan M. Monk<sup>1</sup>, Pieter C. Dorrestein<sup>3,4,7,9</sup>, Joseph Pogliano<sup>2</sup>, Rob Knight<sup>1,5,8,9</sup>, Victor Nizet<sup>4,5,9</sup>, Bernhard O. Palsson<sup>1,5,6,9</sup> & Adam M. Feist<sup>1,6</sup>

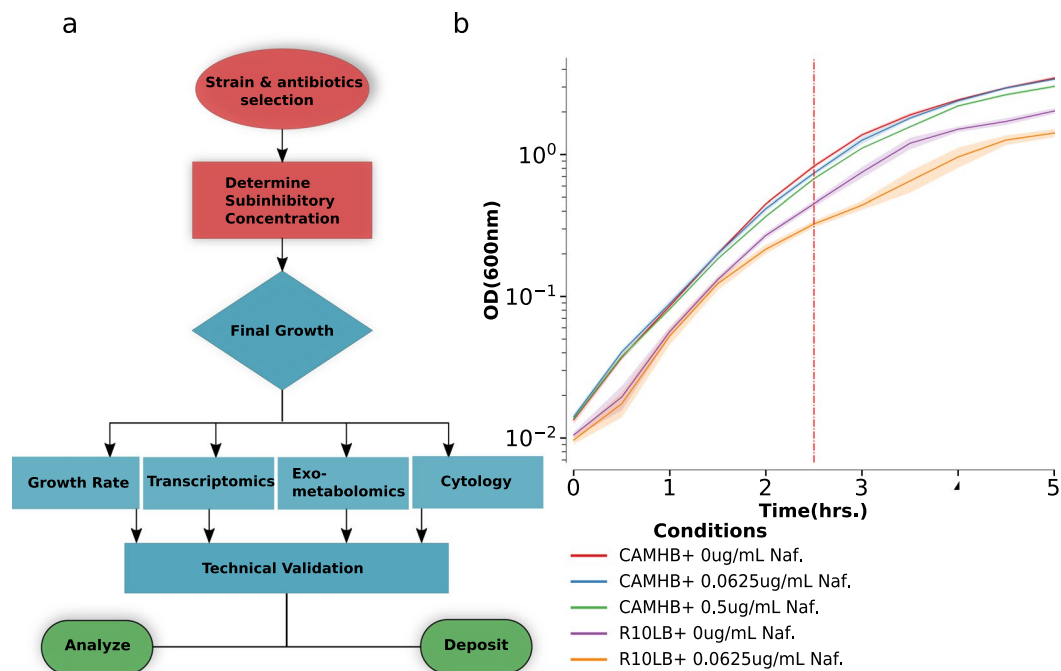
Cation adjusted-Mueller Hinton Broth (CA-MHB) is the standard bacteriological medium utilized in the clinic for the determination of antibiotic susceptibility. However, a growing number of literature has demonstrated that media conditions can cause a substantial difference in the efficacy of antibiotics and antimicrobials. Recent studies have also shown that minimum inhibitory concentration (MIC) tests performed in standard cell culture media (e.g. RPMI and DMEM) are more indicative of *in vivo* antibiotic efficacy, presumably because they are a better proxy for the human host's physiological conditions. The basis for the bacterial media dependent susceptibility to antibiotics remains undefined. To address this question, we characterized the physiological response of methicillin-resistant *Staphylococcus aureus* (MRSA) during exposure to sub-inhibitory concentrations of the beta-lactam antibiotic nafcillin in either CA-MHB or RPMI + 10% LB (R10LB). Here, we present high quality transcriptomic, exo-metabolomic and morphological data paired with growth and susceptibility results for MRSA cultured in either standard bacteriologic or more physiologic relevant medium.

## Background & Summary

The emergence of antimicrobial resistance amongst bacterial pathogens is quickly exhausting our current repertoire of efficacious treatment options. With growing biological and economic limitations on traditional approaches to antibiotic therapy, infectious diseases are on pace to reclaim their status as the leading cause of human mortality worldwide<sup>1</sup>. With the acquisition of multidrug resistance and its rapid spread outside of the nosocomial environment, community-associated methicillin-resistant *S. aureus* (CA-MRSA) strains have exacerbated the burden imposed on modern healthcare<sup>2</sup>. While there have been extensive efforts to develop novel therapeutics we still do not fully understand the bacterial physiological transformations that occur to promote the evolution of antibiotic resistance.

CA-MHB is the standard media used for determining the minimum inhibitory concentration (MIC) of antibiotics against *S. aureus* in both clinical and research settings<sup>3</sup>. It is designed for non-selective and rapid growth

<sup>1</sup>Department of Bioengineering, University of California, San Diego, La Jolla, USA. <sup>2</sup>Division of Biological Sciences, University of California San Diego, La Jolla, CA, 92093, USA. <sup>3</sup>Collaborative Mass Spectrometry Innovation Center, University of California, San Diego, La Jolla, California, USA. <sup>4</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA. <sup>5</sup>Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA. <sup>6</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800, Kongens Lyngby, Denmark. <sup>7</sup>Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, 92093, USA. <sup>8</sup>Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, 92093, USA. <sup>9</sup>Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, 92093, USA. Correspondence and requests for materials should be addressed to A.M.F. (email: [afeist@ucsd.edu](mailto:afeist@ucsd.edu))



**Fig. 1** Data collection. **(a)** Schematic depicting the workflow used to generate the multi-omic data. **(b)** Plot of growth curves during sample collection. Cells were grown for five hours in each condition. At 2.5 hour mark (red dashed line), samples were collected for BCP and Rna seq while samples for exo-metabolomics and growth rate were collected every 30 minutes. The shaded region represents 95% confidence intervals (CI) of the OD600 measurements from three biological replicates.

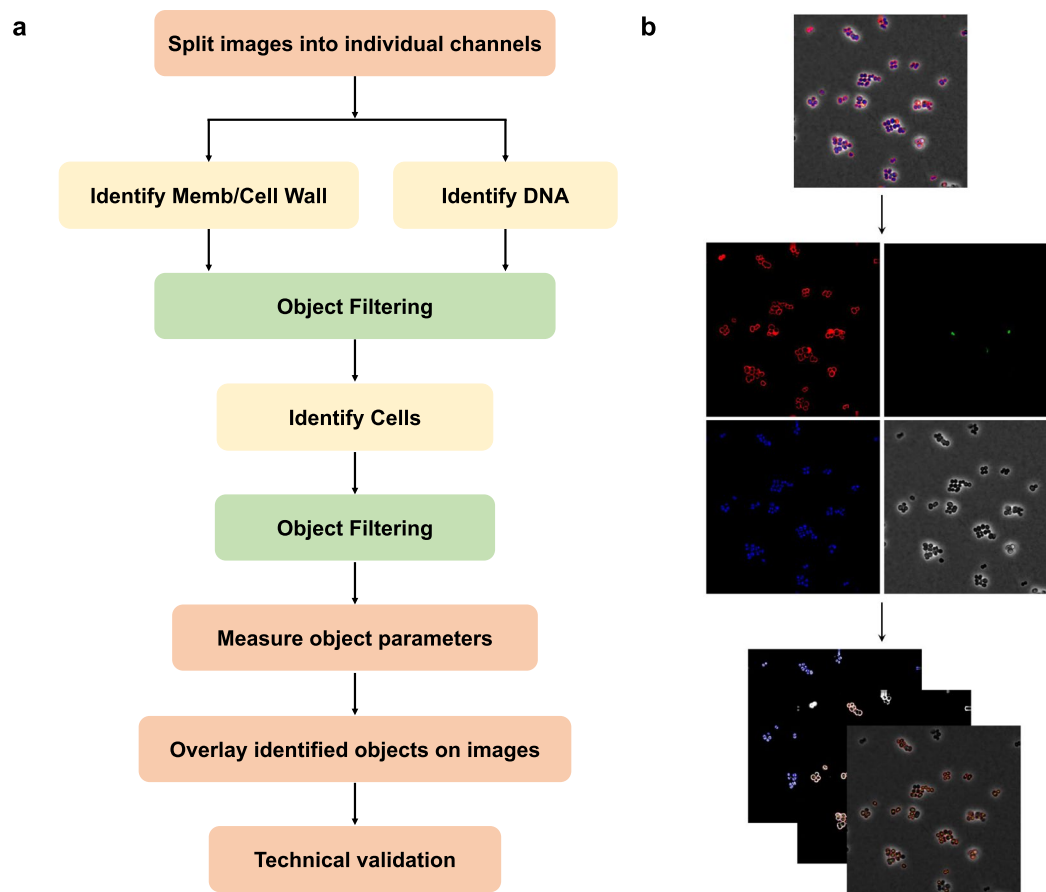
of microbes in ideal conditions. However, many recent studies have shown that a pathogen's response to anti-microbial compounds is dependent on the media conditions in which they were tested and the consequent cellular state<sup>4–6</sup>. Furthermore, Ersoy *et al.* demonstrated that the accuracy of antibiotic susceptibility testing can be improved by using cell culture media that are designed to mimic host physiological conditions, calling into question the prevalent usage of CA-MHB<sup>7</sup>. Though many cellular processes have been linked to changes in MIC, it is still unclear which (if any) of the proposed mechanisms play a role in the differences we observe between bacteriological and physiological media<sup>4,8</sup>.

To address this question, we characterized the response of CA-MRSA isolate TCH1516 to different sub-inhibitory concentrations of nafcillin in CAMHB and RPMI + 10% LB (R10LB) using a workflow designed for multi-omic data collection (Fig. 1a). In accordance with previous findings, we found that the MIC of nafcillin dropped from 12.8 ug/mL in CA-MHB to 0.2 ug/mL in R10LB<sup>7</sup>. Next, we grew the cells in several sub-inhibitory concentrations of nafcillin and chose the conditions that allowed for lowered but continued steady state growth for further analysis. The growth rates under these conditions were checked repeatedly to ensure that they were highly reproducible (Fig. 1b)<sup>9</sup>. This allowed us to streamline our protocol by letting us take samples at the same time point across the two media conditions. Sampling at a consistent time point also had an added advantage of having equal exposure time to nafcillin, which has a time-dependent bactericidal effect<sup>9</sup>. Under these conditions, we characterized TCH1516 response with growth rate measurements, RNA sequencing, exo-metabolomics (HPLC and LC-MS) and Bacterial Cytological Profiling (BCP). This dataset provides multi-omic coverage of cellular response to antibiotic in conditions that lead to drastically different efficacy of the drug. When used in the framework of emerging genome scale models and integration methods, it will provide a more complete understanding of emerging beta-lactam resistance.

## Methods

**Culture and Growth Conditions.** Standard bacteriological media MHB (Sigma-Aldrich) was supplemented with 25 mg/L  $\text{Ca}^{2+}$  and 12.5 mg/L  $\text{Mg}^{2+}$  (CA-MHB). Eukaryotic cell culture media Roswell Park Memorial Institute 1640 (RPMI) (Thermo Fisher Scientific) was supplemented with 10% LB (R10LB). Broth microdilution was performed to determine the nafcillin MIC in each media condition. On the day of the experiment, overnight cultures of CA-MRSA TCH1516 were diluted to a starting OD600 of 0.01 into fresh media and grown at 37°C with stirring to OD600 0.4. This preculture was then diluted back to OD600 0.01 into fresh media containing no drug or sub-inhibitory concentrations of nafcillin relative for each media type. Growth was monitored by obtaining OD600 readings every 30 min for 5 hr. Three biological replicates were collected for the study, each derived from different colony and overnight culture.

**cDNA library preparation and RNA sequencing.** After 2.5 hours of growth, 3 mL samples were taken for RNA sequencing and added to tubes containing 6 mL RNAProtect. After incubation, they were centrifuged



**Fig. 2** Image analysis pipeline for cell morphology. **(a)** Schematic of cytology data pipeline using CellProfiler 3.0. **(b)** Representation of image analysis pipeline. Undeconvolved and deconvolved composite images are split into individual channels. Images were thresholded and filtered WGA and DAPI channels to generate segmentation masks for identified objects.

to remove the supernatant. RNA was extracted from the pelleted cells using a ‘Quick RNA Fungal/Bacterial Microprep’ kit developed by Zymo Research. The cells were mechanically lysed with a Roche MagNa Lyser instrument and DNA was removed with DNase I during the RNA purification. RNA quality was checked with an Agilent Bioanalyzer instrument and ribosomal RNA was removed using an Illumina Ribo-Zero kit. The remaining RNA was used to build a cDNA library for sequencing using a KAPA Stranded RNA-seq Library Preparation Kit. The kit was used for RNA fragmentation, sequencing adapter ligation, and library amplification. The generated cDNA libraries were sent for Illumina sequencing on a HiSeq 4000.

**RNA sequencing analysis.** The phred quality scores for illumina sequencing were generated using Fastqc package<sup>10</sup>. Bowtie2 was used to align the raw reads to TCH1516 genome (NC\_010079.1) and to calculate alignment percentage<sup>11,12</sup>. The aligned reads were then normalized to transcripts per million (TPM) with DESeq2. The ComBat module within the sva package of R was used to correct for batch effects<sup>13,14</sup>. Distance matrix for hierarchical clustering were calculated with sklearn package<sup>15</sup>.

**Bacterial Cytological Profiling.** At the 2.5 hour mark, samples were taken for fluorescence microscopy, similar to previously described with modifications<sup>16</sup>. In brief, 8  $\mu$ L cells were added to 2  $\mu$ L dye mix containing 10  $\mu$ g/mL DAPI, 2.5  $\mu$ M SYTOX Green, and either 20  $\mu$ g/mL WGA-555 or WGA-Rhodamine in 1x T-base. The sample was then transferred to a glass slide containing an agarose pad (20% media, 1.2% agarose) and imaged on an Applied Precision DV Elite epifluorescence microscope with a CMOS camera. The exposure times for each wavelength were as follows, TRITC/Cy-5 = 0.025 s, FITC/FITC = 0.01 s, DAPI/DAPI = 0.015 s, and were kept constant for all images.

Deconvolved images were adjusted using FIJI (ImageJ 1.51w) and Adobe Photoshop (2015.1) to remove background in WGA and DAPI channels and to ensure that cell and DNA objects are within the highest intensity quartile. These images were then processed using a custom CellProfiler 3.0 pipeline that individually thresholded and filtered WGA and DAPI channels to obtain segmentation masks for the cell wall, DNA and entire cell. Objects identified in this manner were further processed in CellProfiler to obtain a total of 5285 features<sup>17–19</sup>. Prior to analysis, feature selection is necessary to create a subset of relevant features as to minimize redundancy within the dataset. The summary of processing steps and an example image are presented in Fig. 2.

**Untargeted Liquid Chromatography Mass Spectrometry Data Acquisition.** Following dilution of the preculture of CA-MRSA TCH1516 into fresh media, approximately 400  $\mu$ L of liquid media containing cells were collected at 30 minutes intervals (at the same time as samples for OD600 measurements) from each of the samples. Growth media was syringe-filtered through 0.22  $\mu$ m disc filters (Millex-GV, MilliporeSigma) to remove cells. The filtered growth media was collected and stored at  $-80^{\circ}\text{C}$  until analysis by liquid chromatography mass spectrometry (LC/MS). For LC/MS analysis, samples were subjected to chromatographic separation using an UltiMate 3000 UHPLC system (Thermo Scientific). Chromatographic separations were achieved using a 50 mm  $\times$  2.1 mm Kinetex 2.6 micron polar-C18 column (Phenomenex) held at a fixed temperature of  $30^{\circ}\text{C}$  within an actively heated column compartment. Samples were injected onto the LC column via thermostatted autosampler maintained at  $4^{\circ}\text{C}$ . For samples containing RPMI + 10% LB media the injection volume was 5  $\mu$ L, while the injection volume was 2  $\mu$ L for samples containing CA-MHB to prevent excessive oversaturation of the mass spectrometer detector due to the higher concentrations of many molecules in the CA-MHB media.

After injection, the sample components were eluted from the LC column into the mass spectrometer using a flow rate of 0.5 mL/min and the following mobile phases: Mobile phase A was LC/MS grade water with 0.1% formic acid (v/v) and mobile phase B was LC/MS grade acetonitrile with 0.1% formic acid (v/v). The LC gradient program was as follows: 0–1.0 min 5%B, 1.0–5.0 min 5–35%B, 5.0–5.5 min 35–100%B, 5.5–6.0 min 100%B, and 6.0–6.5 min 100–5%B followed by 5 minutes of re-equilibration at 5%B. Mass spectrometric data was acquired using a Bruker Daltonics maXis Impact quadrupole-time-of-flight (qTOF) mass spectrometer equipped with an Apollo II electrospray ionization (ESI) source and controlled via otofControl v4.0.15 and Hystar v3.2 software packages (Bruker Daltonics). The mass accuracy of the maXis instrument was first externally calibrated using a calibration solution of sodium formate which provided  $>21$  reference  $m/z$ 's between 50–1500  $m/z$  of the mass spectrum in both positive and negative polarities (reference  $m/z$  list provided within instrument control software). Sodium formate solution was prepared using 9.9 ml of 50/50% isopropanol/water, plus 0.2% formic acid, and 100  $\mu$ L of 1 M NaOH. During infusion of all samples, the mass accuracy of the instrument was maintained to  $<10$  ppm via constant introduction of and internal calibrant, or “lock mass”, in the form of hexakis (1H,1H,2H-difluoroethoxy)phosphazene (SynQuest Labs, Inc.). During positive polarity runs the lock mass compound was detected as the ion  $m/z$  622.028960 ( $\text{C}_{12}\text{H}_{19}\text{F}_{12}\text{N}_3\text{O}_6\text{P}_3^+$ ) and in negative polarity the lock mass compound formed the ion  $m/z$  556.001951 ( $\text{C}_{10}\text{H}_{15}\text{F}_{10}\text{N}_3\text{O}_6\text{P}_3^-$ ).

Instrument source parameters were set as follows: nebulizer gas (Nitrogen) pressure, 2 Bar; Capillary voltage, 3,500 V; ion source temperature,  $200^{\circ}\text{C}$ ; dry gas flow, 9 L/min. The global mass spectral acquisition rate was set at 3 Hz. The instrument transfer optics were tuned as follows: Ion funnel 1 & 2 RFs of 250 Vpp (volts peak-to-peak), hexapole RF of 100 Vpp, quadrupole ion energy of 5 eV, collision quadrupole energy of 5 eV, and a TOF pre-pulse storage of 7.0  $\mu$ secs. The post-collision quadrupole RF and TOF transfer time were stepped across four values per MS scan. The collision RF was stepped at 450, 550, 800, 1100 Vpp. The transfer time was stepped at 70, 75, 90, 95  $\mu$ secs. All samples were run twice, once under positive polarity settings and once under negative polarity settings.

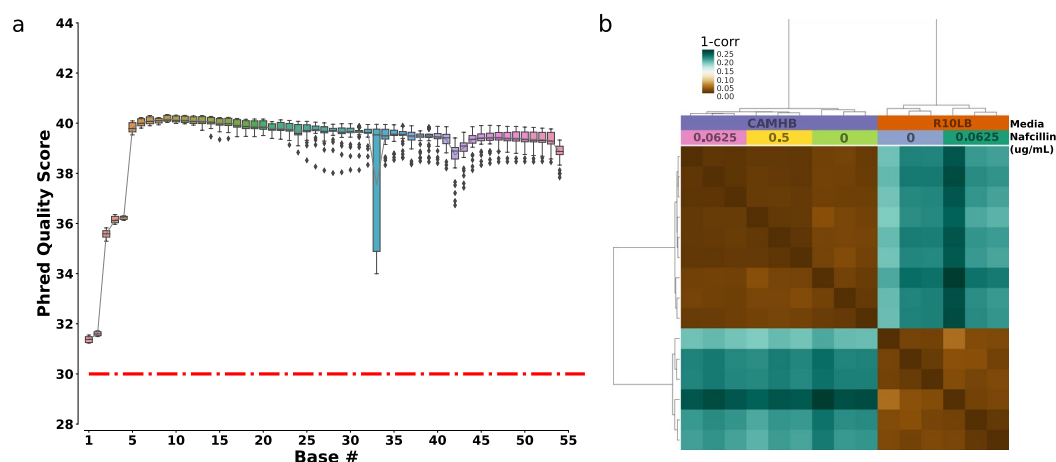
Following acquisition of the LC/MS data, lock mass calibration was applied to all data files in order to apply a linear correction calibration to all  $m/z$  values recorded in each mass spectrum. The application of this mass correction was applied automatically via the Bruker Daltonics Compass Data Analysis software (ver. 4.3.110), using the  $m/z$  of the hexakis (1H,1H,2H-difluoroethoxy) phosphazene as the reference lock mass calibration compound. Following lock mass re-calibration of the data, all files were converted from the Bruker Daltonics proprietary format (.d) and exported to an open data format known as .mzXML. All data herein was deposited to MassIVE<sup>20</sup>.

**Targeted High-Performance Liquid Chromatography.** For organic acid and carbohydrate detection, samples were collected every 30 minutes and filtered as described above. The filtered samples were loaded onto a 1260 Infinity series (Agilent Technologies) high-performance liquid chromatography (HPLC) system with an Aminex HPX-87H column (Bio-Rad Laboratories) and a refractive index detector. The system was operated using ChemStation software. The HPLC was run with a single mobile phase composed of HPLC grade water buffered with 5 mM sulfuric acid ( $\text{H}_2\text{SO}_4$ ). The flow rate was held at 0.5 mL/minute, the sample injection volume was 10  $\mu$ L, and the column temperature was maintained at  $45^{\circ}\text{C}$ . The identities of compounds were determined by retention time comparison to standard curves of acetate, ethanol, glucose, lactate, pyruvate, and succinate. The peak area integration and resulting chromatograms were generated within ChemStation and compared to that of the standard curves in order to determine the concentration of each compound in the samples. These final concentration values were deposited to MassIVE database<sup>20</sup>.

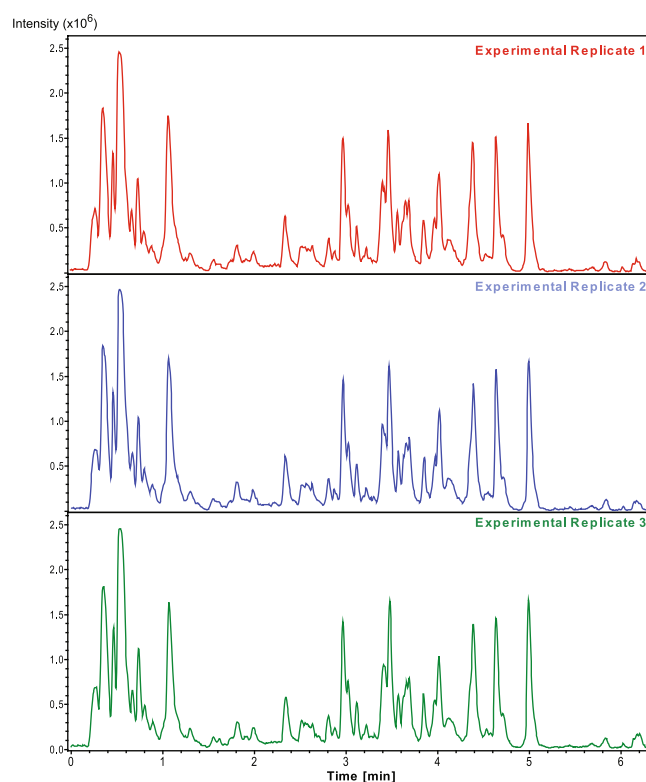
**Exclusion Criteria.** Some samples with low concentrations of antibiotics were excluded due to lack of statistically significant change in measured phenotypes. These measurements are available in public databases under same project numbers as those included in this manuscript or upon request.

### Data Records

The growth-rate data has been deposited in figshare<sup>9</sup>. HPLC, mass spectrometry as well as the cytological profiling data can be found in MassIVE<sup>20</sup>. The pipeline used to process and quality check RNA sequencing data has been deposited in figshare<sup>21,22</sup>. Raw FASTQ sequencing files for RNA sequencing have been deposited to the NCBI Sequence Read Archive<sup>23</sup>. The processed RNA sequencing file containing raw and ComBat corrected TPM and counts were deposited in Gene Expression Omnibus<sup>24</sup>. The summary statistics from RNA sequencing pipeline was compiled and submitted to figshare<sup>25</sup>.



**Fig. 3** RNA sequencing technical validation. **(a)** A distribution plot of base wise phred quality score of all samples in the dataset. The red dashed line at the score of 30 corresponds to 99.9% accuracy in base calling. **(b)** Clustering of samples based on Spearman correlation of sample TPMs.



**Fig. 4** Validation of LC/MS base peak chromatograms of experimental replicates. **(a)** Evaluation of the reproducibility of both retention time and peak areas within the LC/MS data could be performed by direct comparison of the base peak chromatograms of experimental replicates. The plot represents a comparison of three separate experimental replicates of the same time-point from the same type of sample, in this case untreated TCH1516 grown in RPMI + 10% LB at 1 hr.

## Technical Validation

**RNA sequencing.** Each library was run on an Agilent Technologies TapeStation to check for correct size distribution and for the removal of unincorporated primers and adapters. After pooling libraries with distinct barcodes, a final size selection was performed with SPRI beads to remove sub-detectable PCR primer carryover. This final pool was checked again on the TapeStation prior to being run on the Illumina HiSeq 4000.

The quality of sequenced reads, as measured by average Phred Scores, was greater than 30 across all samples (Fig. 3a). These scores correspond to base calling accuracy of at least 99.9%, though for most bases the scores



were much higher. These raw fastq files were deposited to SRA database<sup>23</sup>. For additional quality checks, the reads were next aligned to the TCH1516 genome (NC\_010079.1) with Bowtie2. For each sample, at least 90% of reads successfully aligned to the genome<sup>25</sup>. Due to large number of collected samples, they were processed in 3 different batches. Each batch contained a full set of samples (i.e. all media types and all antibiotic concentrations) which allowed us to correct for any batch effect in the dataset. As a demonstration, we used the ComBat module within SVA package of R to correct for batch effect which resulted in highly reproducible data (Spearman  $r > 0.98$ ) with close clustering of all biological replicates (Fig. 3b). The ComBat batch corrected TPM were deposited into GEO database<sup>24</sup>. As with ComBat, most modern differential expression analysis packages such as DESeq2 and EdgeR have built in ways to model these covariates.

**Bacterial Cytological Profiling.** Prior to processing with CellProfiler, images must be manually screened for issues that would cause errors during image segmentation. Image adjustments such as contrast and brightness as well as background removal are used to reduce segmentation errors.

Following CellProfiler analysis, images of each object segmentation (13 in total) are generated and visually screened to ensure that measurements are obtained from the correct object. Additionally, prior to analysis, parameters from cell outlines must be combined with their related objects such as DNA. For each, a reference number is generated at the time of measurement and must match in both the cell and DNA objects ensuring that no cell-DNA mismatches are possible. A second checkpoint exists in the form of the “parent” tag for DNA objects in which the number of unique tags must match the number of cell outlines for that corresponding image. Excel files containing cellular features were uploaded to MassIVE<sup>20</sup>.

**Untargeted Liquid Chromatography Mass Spectrometry Data Acquisition.** Global retention time reproducibility and ion intensity reproducibility was evaluated across all samples by direct comparison of both the base peak chromatogram (BPCs) and multiple extracted ion chromatograms (EICs) from each sample. Comparison of BPCs between experimental triplicate of each sample measured give an overall representation of the reproducibility of both the retention time and the peak intensity (Fig. 4). Additionally, the EICs for the various molecules determined to be present in each media could be evaluated across replicates to determine the more precise retention time drift ( $<0.1$  minutes) and peak area (coefficient of variation  $<15\%$ ).

## Code Availability

The pipeline used to analyze RNA sequence data is available on figshare<sup>21</sup>.

The script used to remove batch effects from RNA sequencing data is available on figshare<sup>22</sup>.

## References

- O'Neill, J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations* (2016).
- David, M. Z. & Daum, R. S. Community-associated methicillin-resistant *Staphylococcus aureus*: epidemiology and clinical consequences of an emerging epidemic. *Clin. Microbiol. Rev.* **23**, 616–687 (2010).
- Clinical and Laboratory Standards Institute, *Performance Standards for Antimicrobial Susceptibility Testing*. (CLSI, 2014).
- Martinez, J. L. & Rojo, F. Metabolic regulation of antibiotic resistance. *FEMS Microbiol. Rev.* **35**, 768–789 (2011).
- De Jonge, B. L., Chang, Y. S., Xu, N. & Gage, D. Effect of exogenous glycine on peptidoglycan composition and resistance in a methicillin-resistant *Staphylococcus aureus* strain. *Antimicrob. Agents Chemother.* **40**, 1498–1503 (1996).
- Dorschner, R. A. *et al.* The mammalian ionic environment dictates microbial susceptibility to antimicrobial defense peptides. *The FASEB Journal* **20**, 35–42 (2006).
- Ersoy, S. C. *et al.* Correcting a Fundamental Flaw in the Paradigm for Antimicrobial Susceptibility Testing. *E Bio Medicine* **20**, 173–181 (2017).
- Poole, K. Bacterial stress responses as determinants of antimicrobial resistance. *J. Antimicrob. Chemother.* **67**, 2069–2089 (2012).
- Poudel, S. OD600 Data. *figshare* <https://doi.org/10.6084/m9.figshare.7627451> (2018).
- Tournidge, J. D. The pharmacodynamics of  $\beta$ -lactams. *Clin. Infect. Dis.* **27**, 10–22 (1998).
- Andrews, S. A quality control tool for high throughput sequence data. *FastQC*, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
- Highlander, S. K. *et al.* Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiol.* **7**, 99 (2007).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
- Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
- Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- Quach, D. T., Sakoulas, G., Nizet, V., Pogliano, J. & Pogliano, K. Bacterial Cytological Profiling (BCP) as a Rapid and Accurate Antimicrobial Susceptibility Testing Method for *Staphylococcus aureus*. *E Bio Medicine* **4**, 95–103 (2016).
- Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
- Rodenacker, K. & Bengtsson, E. A feature set for cytometry on digitized microscopic images. *Anal. Cell. Pathol.* **25**, 1–36 (2003).
- Palsson, B. *Staphylococcus aureus* TCH1516 LC/MS, HPLC, BPC data. *MassIVE*, <https://identifiers.org/massive:MSV000082804> (2018).
- Poudel, S. SciData\_RNAseq\_Workflow. *figshare* <https://doi.org/10.6084/m9.figshare.7658591> (2018).
- Poudel, S. TCH1516\_BatchCorrection\_SciData. *figshare* <https://doi.org/10.6084/m9.figshare.7603388> (2018).
- NCBI Sequence Read Archive <http://identifiers.org/ncbi/insdc.sra:SRP127172> (2018).
- Gene Expression Omnibus <https://identifiers.org/GEO:GSE115309> (2018).
- Poudel, S. RNAseq Summary Statistics. *figshare* <https://doi.org/10.6084/m9.figshare.7624949> (2018).

## Acknowledgements

We thank Anand Sastry for helping build the RNA sequencing analysis pipeline. This research was supported by NIH NIAID grant (1-U01-AI124316).

### Author Contributions

S.P. analyzed RNA sequencing data and wrote Data Descriptor, Methods and Technical Validation. H.T. performed growth experiments, analyzed BCP data, and wrote Methods and Technical Validation. M.M. analyzed HPLC data and wrote Data Descriptor, Methods and Technical Validation. R.S. prepared samples for RNA sequencing and wrote Methods and Technical Validation. C.A.O. prepared samples for HPLC and wrote Methods and Technical Validation. A.L. performed growth experiments, analyzed BCP data, and wrote Methods and Technical Validation. Y.S. analysed HPLC data. N.D. performed preliminary growth and MIC experiments. A.V. performed growth experiments. J.S. wrote Technical Validation and Methods. S.M.D. performed growth experiments.

### Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019